

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:

- (A) NAME: PLANT BIOSCIENCE LIMITED
 (B) STREET: COLNEY LANE
 (C) CITY: NORWICH
 (D) STATE: NORFOLK
 (E) COUNTRY: UK
 (F) POSTAL CODE (ZIP): NR4 7UH

(ii) TITLE OF INVENTION: PROTEIN

(iii) NUMBER OF SEQUENCES: 4

(iv) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Floppy disk
 (B) COMPUTER: IBM PC compatible
 (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 (D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1205 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

- (A) NAME/KEY: CDS
 (B) LOCATION: 18..917

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

CCAAATCCCA TCTCAAA ATG AAT GAT CTA TTT TCA GGA TCT TTC TCT CGT	50
Met Asn Asp Leu Phe Ser Gly Ser Phe Ser Arg	
1 5 10	
TTC AGA GCT GAC GAT CAA TCG GAC TCT CAC GCC ATA GAA ATG GGA GAC	98
Phe Arg Ala Asp Asp Gln Ser Asp Ser His Ala Ile Glu Met Gly Asp	
15 20 25	
ATT ACT GGC GGA GTC AAT CTC GAC AAA TTC TTC GAA GAT GTT GAA GCC	146
Ile Thr Gly Gly Val Asn Leu Asp Lys Phe Phe Glu Asp Val Glu Ala	
30 35 40	
ATT AAA GAC GAA CTC AAA GGC CTC GAG AAA ATC TAT TCC CAA CTC CAA	194
Ile Lys Asp Glu Leu Lys Gly Leu Glu Lys Ile Tyr Ser Gln Leu Gln	
45 50 55	
TCT TCC CAT GAA AAA AGC AAG ACT CTT CAC AAC GCT AAA GCC GTT AAA	242
Ser Ser His Glu Lys Ser Lys Thr Leu His Asn Ala Lys Ala Val Lys	
60 65 70 75	
GAT CTA AGA TCC AAC ATG GAT AAT GAC GTT TCC ATG GCA TTG AAG AAA	290
Asp Leu Arg Ser Asn Met Asp Asn Asp Val Ser Met Ala Leu Lys Lys	
80 85 90	
GCC AAA TTC ATC AAA GTT CGT CTC GAA GCC TTA GAC AGA TCA AAT GCA	338

Ala Lys Phe Ile Lys Val Arg Leu Glu Ala Leu Asp Arg Ser Asn Ala	
95 100 105	
GCG AAT CGA AGC CTC CCT GGA TGT GGA CCC GGA AGT TCA TCT GAC AGG	386
Ala Asn Arg Ser Leu Pro Gly Cys Gly Pro Gly Ser Ser Ser Asp Arg	
110 115 120	
ACG AGA ACT TCA GTT GTG AAC GGA TTA AGG AAG AAA CTT CAA GAG TCA	434
Thr Arg Thr Ser Val Val Asn Gly Leu Arg Lys Lys Leu Gln Glu Ser	
125 130 135	
ATG AAT CAG TTC AAC GAG CTA AGG CAA AAG ATG GCA TCT GAA TAT AGG	482
Met Asn Gln Phe Asn Glu Leu Arg Gln Lys Met Ala Ser Glu Tyr Arg	
140 145 150 155	
GAA ACA GTT CAA CGA CGA TAT TAT ACC GTC ACA GGA GAA AAT CCT GAT	530
Glu Thr Val Gln Arg Arg Tyr Tyr Thr Val Thr Gly Glu Asn Pro Asp	
160 165 170	
GAA GCA GTT CTT GAT ACA CTC ATA TCT ACA GGT CAA AGT GAG ACG TTC	578
Glu Ala Val Leu Asp Thr Leu Ile Ser Thr Gly Gln Ser Glu Thr Phe	
175 180 185	
TTG CAA AAG GCA ATT CAA GAG CAA GGG AGA GGA CAA GTG ATG GAT ACA	626
Leu Gln Lys Ala Ile Gln Glu Gln Gly Arg Gly Gln Val Met Asp Thr	
190 195 200	
GTT ATG GAA ATT CAA GAA AGG CAT GAA GCT GTG AAG GAA TTG GAG AGG	674
Val Met Glu Ile Gln Glu Arg His Glu Ala Val Lys Glu Leu Glu Arg	
205 210 215	
AAT TTG AAA GAA TTG CAT CAA GTA TTC TTG GAC ATG GCT GTT TTG GTT	722
Asn Leu Lys Glu Leu His Gln Val Phe Leu Asp Met Ala Val Leu Val	
220 225 230 235	
GAA AGT CAA GGA GCT CAA CTT GAT GAT ATT GAG AGC CAA GTG AAT AGG	770
Glu Ser Gln Gly Ala Gln Leu Asp Asp Ile Glu Ser Gln Val Asn Arg	
240 245 250	
GCT AAT TCC TTC GTT AGA GGG GGT GCT CAG CAA CTG CAA GTG GCA AGG	818
Ala Asn Ser Phe Val Arg Gly Gly Ala Gln Gln Leu Gln Val Ala Arg	
255 260 265	
AAG CAC CAG AAG AAC ACT AGA AAA TGG ACT TGT TTT GCT ATT ATT CTT	866
Lys His Gln Lys Asn Thr Arg Lys Trp Thr Cys Phe Ala Ile Ile Leu	
270 275 280	
CTG CTT ATC ATC ATT TTG GTG GTG GTT CTT TCT ATT CAG CCA TGG AAA	914
Leu Leu Ile Ile Ile Leu Val Val Val Leu Ser Ile Gln Pro Trp Lys	
285 290 295	
AAA TGAGAATTG TCTATGGTCA AAGGTCTTCT GGTGGACCCC TTCAATGTTT	967
Lys	
300	
TGAATATTCT AAATTTTAT ATTTTATTAT TTTAGCCATG CTTATTATTT TGTGTTATTT	1027
TGGATTTTTT TTTTGTTTTT AATGTGGGGA AGAGTAACT GGATGGGGGT CCATGTGCTA	1087
TTTAGAGAAA TACTTGGGAG TTCTCTTTTT GTAATTATTG CTGTATTAG AGTATAATTC	1147
TTTTTCTATA TGTGGGCAG GTTAATTTGT TTGTTTGATT ATATTCTCAT TTAGATTT	1205

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 300 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```

Met Asn Asp Leu Phe Ser Gly Ser Phe Ser Arg Phe Arg Ala Asp Asp
 1           5           10          15
Gln Ser Asp Ser His Ala Ile Glu Met Gly Asp Ile Thr Gly Gly Val
 20          25          30
Asn Leu Asp Lys Phe Phe Glu Asp Val Glu Ala Ile Lys Asp Glu Leu
 35          40          45
Lys Gly Leu Glu Lys Ile Tyr Ser Gln Leu Gln Ser Ser His Glu Lys
 50          55          60
Ser Lys Thr Leu His Asn Ala Lys Ala Val Lys Asp Leu Arg Ser Asn
 65          70          75          80
Met Asp Asn Asp Val Ser Met Ala Leu Lys Lys Ala Lys Phe Ile Lys
 85          90          95
Val Arg Leu Glu Ala Leu Asp Arg Ser Asn Ala Ala Asn Arg Ser Leu
100          105          110
Pro Gly Cys Gly Pro Gly Ser Ser Ser Asp Arg Thr Arg Thr Ser Val
115          120          125
Val Asn Gly Leu Arg Lys Lys Leu Gln Glu Ser Met Asn Gln Phe Asn
130          135          140
Glu Leu Arg Gln Lys Met Ala Ser Glu Tyr Arg Glu Thr Val Gln Arg
145          150          155          160
Arg Tyr Tyr Thr Val Thr Gly Glu Asn Pro Asp Glu Ala Val Leu Asp
165          170          175
Thr Leu Ile Ser Thr Gly Gln Ser Glu Thr Phe Leu Gln Lys Ala Ile
180          185          190
Gln Glu Gln Gly Arg Gly Gln Val Met Asp Thr Val Met Glu Ile Gln
195          200          205
Glu Arg His Glu Ala Val Lys Glu Leu Glu Arg Asn Leu Lys Glu Leu
210          215          220
His Gln Val Phe Leu Asp Met Ala Val Leu Val Glu Ser Gln Gly Ala
225          230          235          240
Gln Leu Asp Asp Ile Glu Ser Gln Val Asn Arg Ala Asn Ser Phe Val
245          250          255
Arg Gly Gly Ala Gln Gln Leu Gln Val Ala Arg Lys His Gln Lys Asn
260          265          270
Thr Arg Lys Trp Thr Cys Phe Ala Ile Ile Leu Leu Leu Ile Ile Ile
275          280          285
Leu Val Val Val Leu Ser Ile Gln Pro Trp Lys Lys
290          295          300

```

(2) INFORMATION FOR SEQ ID NO: 3:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1334 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

- (ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 77..991

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

```

GAATTCCTCG AGCTACGTCA GGGATTCATT CCGATCTGAA ATCTCTCTCT AGATTCTCT      60
ATTTTTCGAA TTTTAA ATG AAC GAT TTG TTT TCC AGC TCA TTC TCT CGC      109
      Met Asn Asp Leu Phe Ser Ser Ser Phe Ser Arg
              1              5              10

TTC CGC AGC GGA GAA CCA TCC CCT CGC CGA GAC GTT GCC GGC GGT GGC      157
Phe Arg Ser Gly Glu Pro Ser Pro Arg Arg Asp Val Ala Gly Gly Gly
              15              20              25

GAC GGA GTT CAG ATG GCG AAT CCC GCG GGA TCA ACC GGT GGT GTG AAC      205
Asp Gly Val Gln Met Ala Asn Pro Ala Gly Ser Thr Gly Gly Val Asn
              30              35              40

CTC GAC AAG TTC TTC GAA GAT GTT GAA TCT GTG AAA GAA GAG CTA AAG      253
Leu Asp Lys Phe Phe Glu Asp Val Glu Ser Val Lys Glu Glu Leu Lys
              45              50              55

GAG CTA GAT CGG CTC AAC GAA ACA CTC TCT TCA TGT CAC GAG CAG AGC      301
Glu Leu Asp Arg Leu Asn Glu Thr Leu Ser Ser Cys His Glu Gln Ser
              60              65              70              75

AAG ACG CTT CAC AAT GCT AAA GCC GTT AAA GAT CTC CGG TCT AAA ATG      349
Lys Thr Leu His Asn Ala Lys Ala Val Lys Asp Leu Arg Ser Lys Met
              80              85              90

GAC GGT GAC GTT GGA GTC GCG TTG AAG AAG GCG AAG ATG ATT AAA GTT      397
Asp Gly Asp Val Gly Val Ala Leu Lys Lys Ala Lys Met Ile Lys Val
              95              100              105

AAA CTC GAG GCG CTA GAT CGT GCC AAT GCT GCT AAT CGG AGT CTC CCT      445
Lys Leu Glu Ala Leu Asp Arg Ala Asn Ala Ala Asn Arg Ser Leu Pro
              110              115              120

GGC TGT GGA CCT GGT TCT TCC TCC GAT CGA ACC AGG ACC TCT GTC CTC      493
Gly Cys Gly Pro Gly Ser Ser Ser Asp Arg Thr Arg Thr Ser Val Leu
              125              130              135

AAT GGT CTC AGG AAG AAA TTG ATG GAC TCT ATG GAT AGT TTC AAC CGA      541
Asn Gly Leu Arg Lys Lys Leu Met Asp Ser Met Asp Ser Phe Asn Arg
              140              145              150              155

TTG AGG GAG CTT ATC TCG TCC GAG TAT AGA GAA ACT GTA CAG AGG AGG      589
Leu Arg Glu Leu Ile Ser Ser Glu Tyr Arg Glu Thr Val Gln Arg Arg
              160              165              170

TAC TTC ACC GTC ACC GGC GAG AAT CCG GAT GAA CGA ACC CTA GAT CGA      637
Tyr Phe Thr Val Thr Gly Glu Asn Pro Asp Glu Arg Thr Leu Asp Arg
              175              180              185

CTG ATT TCC ACT GGA GAG AGT GAG AGA TTC TTG CAG AAA GCA ATA CAA      685

```

Leu Ile Ser Thr Gly Glu Ser Glu Arg Phe Leu Gln Lys Ala Ile Gln	
190 195 200	
GAA CAA GGA AGA GGA AGG GTG TTA GAC ACC ATT AAC GAG ATT CAA GAA	733
Glu Gln Gly Arg Gly Arg Val Leu Asp Thr Ile Asn Glu Ile Gln Glu	
205 210 215	
AGG CAT GAT GCG GTT AAA GAC ATT GAG AAG AAT CTC AGG GAG CTT CAC	781
Arg His Asp Ala Val Lys Asp Ile Glu Lys Asn Leu Arg Glu Leu His	
220 225 230 235	
CAG GTG TTT CTA GAC ATG GCC GTG CTG GTA GAG CAC CAG GGA GCT CAG	829
Gln Val Phe Leu Asp Met Ala Val Leu Val Glu His Gln Gly Ala Gln	
240 245 250	
CTT GAT GAC ATC GAG AGT CAT GTG GGT CGA GCT AGC TCC TTT ATC AGA	877
Leu Asp Asp Ile Glu Ser His Val Gly Arg Ala Ser Ser Phe Ile Arg	
255 260 265	
GGC GGA ACT GAC CAG CTA CAA ACC GCT CGG GTT TAC CAG AAG AAC ACG	925
Gly Gly Thr Asp Gln Leu Gln Thr Ala Arg Val Tyr Gln Lys Asn Thr	
270 275 280	
CGA AAA TGG ACA TGT ATT GCC ATT ATT ATT CTC ATC ATC ATC ATA ACT	973
Arg Lys Trp Thr Cys Ile Ala Ile Ile Ile Leu Ile Ile Ile Thr	
285 290 295	
GTT GTG GTT CTT GCT GTT TTAACCGT GGAACAACAG CAGTGGCGGC	1021
Val Val Val Leu Ala Val	
300 305	
GGCGGCGGTG GTGGTGGTGG GGGTACCACT GGAGGAAGTC AACCAAATTC AGGGACACCA	1081
CCAAATCCTC CTCAGGCAAG GCGTCTATTG CGTTGAAGTT GAAGTTGAAG TTGAGTTTCG	1141
TTATTTCAT ATATATTCTT TCTTTGAAAA ACCTTATTAT CAAACCAGCT TTGTGTTACT	1201
ACTTTCTACT GCTGGTTTGT TGTTAATCTC CCGTTTATTT GGTTTTGTG AAAGAATTTA	1261
AAATGTGGGT TAGATGAGAA AATTAGTACA ACATTCTCTT GTATCTATGT TTGCTACCCT	1321
GACGTAGCTC GAG	1334

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 305 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Met Asn Asp Leu Phe Ser Ser Ser Phe Ser Arg Phe Arg Ser Gly Glu
1 5 10 15
Pro Ser Pro Arg Arg Asp Val Ala Gly Gly Gly Asp Gly Val Gln Met
20 25 30
Ala Asn Pro Ala Gly Ser Thr Gly Gly Val Asn Leu Asp Lys Phe Phe
35 40 45
Glu Asp Val Glu Ser Val Lys Glu Glu Leu Lys Glu Leu Asp Arg Leu
50 55 60

Asn Glu Thr Leu Ser Ser Cys His Glu Gln Ser Lys Thr Leu His Asn
 65 70 75 80
 Ala Lys Ala Val Lys Asp Leu Arg Ser Lys Met Asp Gly Asp Val Gly
 85 90 95
 Val Ala Leu Lys Lys Ala Lys Met Ile Lys Val Lys Leu Glu Ala Leu
 100 105 110
 Asp Arg Ala Asn Ala Ala Asn Arg Ser Leu Pro Gly Cys Gly Pro Gly
 115 120 125
 Ser Ser Ser Asp Arg Thr Arg Thr Ser Val Leu Asn Gly Leu Arg Lys
 130 135 140
 Lys Leu Met Asp Ser Met Asp Ser Phe Asn Arg Leu Arg Glu Leu Ile
 145 150 155 160
 Ser Ser Glu Tyr Arg Glu Thr Val Gln Arg Arg Tyr Phe Thr Val Thr
 165 170 175
 Gly Glu Asn Pro Asp Glu Arg Thr Leu Asp Arg Leu Ile Ser Thr Gly
 180 185 190
 Glu Ser Glu Arg Phe Leu Gln Lys Ala Ile Gln Glu Gln Gly Arg Gly
 195 200 205
 Arg Val Leu Asp Thr Ile Asn Glu Ile Gln Glu Arg His Asp Ala Val
 210 215 220
 Lys Asp Ile Glu Lys Asn Leu Arg Glu Leu His Gln Val Phe Leu Asp
 225 230 235 240
 Met Ala Val Leu Val Glu His Gln Gly Ala Gln Leu Asp Asp Ile Glu
 245 250 255
 Ser His Val Gly Arg Ala Ser Ser Phe Ile Arg Gly Gly Thr Asp Gln
 260 265 270
 Leu Gln Thr Ala Arg Val Tyr Gln Lys Asn Thr Arg Lys Trp Thr Cys
 275 280 285
 Ile Ala Ile Ile Ile Leu Ile Ile Ile Ile Thr Val Val Val Leu Ala
 290 295 300
 Val
 305

SEQ. ID. No. 5

TTTAGATTACTCTTATATTAGTTTGTTTGTAAATTGGACGGTTGTTATATCTTTTTCTTA
ATATGAGATTTATGTCGTTATTAATGTTTTCTCTTGAGGGTTCATAAAGAGATTTATCGTG
TACCTGGGGGTAGGTCAAATGAGAAGGGGTGTAATTTTTGTTTTTTTTTAGGTTTTATTGT
GTTTTATTATTCGTACCGATTTTATTATTTTATATTTTAAATCTTATAAGTTTTGTAAGTT
CCCCAGGTGGTCTTCTGGAACTGGTATCTGTTAAGAGTAAAAAGGTACCGACTTATCTT
TCTTGGTGGTGGTTTTACTACTATTCTGTTCTTATTATCGTTTTGTTTCAGGTAAAAGATCA
CAAGAAGACCACGAAGGAACGGTGAACGTCAACGACTCGTGGGGGAGATTGCTTCCTTAATC
GGGATAAGTGAACCGAGAGTTATAGTAGTTCAACTCGAGGAACTGAAAGTTGGTTTTGTCGG
TACAGGTTCTTATGAACTACGTTAAGAAAGTTTAAGGAGAGGTTAAGGAAGTGTCGAAGTAC
GGAAAGAACTTAAAGGTATTGACATAGGTAGTGAACAGGAGAGGGAACGAGAACTTAACGGA
AAACGTTCTTGACAGAGTGAACTGGACATCTATACTCACATAGTTCTTGACGAAGTAGTCCT
AAAAGAGGACACTGCCATATTATAGCAGCAACTTGACAAAGGGATATAAGTCTACGGTAGAA
AACGGAATCGAGCAACTTGACTAAGTAACTGAGAACTTCAAAGAAGGAATTAGGCAAGTGTT
GACTTCAAGAGCAGGACAGTCTACTTGAAGGCCAGGTGTAGGTCCCTCCGAAGCTAAGCGA
CGTAAACTAGACAGATTCCGAAGCTCTGCTTGAACTACTTAAACCGAAAGAAGTTACGGTA
CCTTTGCAGTAATAGGTACAACCTAGAATCTAGAAATTGCCGAAATCGCAACACTTCTCAGA
ACGAAAAAAGTACCCTTCTAACCTCAACCCTTATCTAAAAGAGCTCCGGAACTCAAGCAGA
AATTACCGAAGTTGTAGAAGCTTCTTAAACAGCTCTAACTGAGGCGGTCATTACAGAGGGTA
AAGATACCGCACTCTCAGGCTAACTAGCAGTCGAGACTTTGCTCTCTTCTAGGACTTTTAT
CTAGTAAGTAAACTCTACCCTAAACC